

10/529263

PCT/JP03/12252

25.09.03

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日  
Date of Application: 2002年 9月25日

REC'D 13 NOV 2003

出 願 番 号  
Application Number: 特願2002-279513  
[ST. 10/C]: [JP2002-279513]

WIPO

PCT

出 願 人  
Applicant(s):

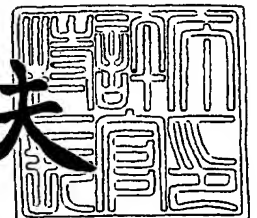
株式会社ベネッセコーポレーション  
藤森 進

PRIORITY DOCUMENT  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)

2003年10月30日

特許庁長官  
Commissioner,  
Japan Patent Office

今井康夫



【書類名】 特許願

【整理番号】 022032

【提出日】 平成14年 9月25日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 15/74

【発明者】

    【住所又は居所】 東京都多摩市落合1丁目34番地 株式会社ベネッセコーポレーション内

    【氏名】 山下 仁司

【発明者】

    【住所又は居所】 東京都府中市武蔵台2-36-10

    【氏名】 藤森 進

【特許出願人】

    【識別番号】 591121498

    【氏名又は名称】 株式会社ベネッセコーポレーション

【特許出願人】

    【住所又は居所】 東京都府中市武蔵台2-36-10

    【氏名又は名称】 藤森 進

【代理人】

    【識別番号】 100089705

    【住所又は居所】 東京都千代田区大手町二丁目2番1号 新大手町ビル206区 ユアサハラ法律特許事務所

    【弁理士】

    【氏名又は名称】 社本 一夫

    【電話番号】 03-3270-6641

【選任した代理人】

    【識別番号】 100076691

    【弁理士】

    【氏名又は名称】 増井 忠次

## 【選任した代理人】

【識別番号】 100075270

【弁理士】

【氏名又は名称】 小林 泰

## 【選任した代理人】

【識別番号】 100080137

【弁理士】

【氏名又は名称】 千葉 昭男

## 【選任した代理人】

【識別番号】 100096013

【弁理士】

【氏名又は名称】 富田 博行

## 【選任した代理人】

【識別番号】 100087424

【弁理士】

【氏名又は名称】 大塚 就彦

## 【手数料の表示】

【予納台帳番号】 051806

【納付金額】 21,000円

## 【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 部分得点モデルに基づくコンピュータ適応型テスト制御システム及び方法

【特許請求の範囲】

【請求項1】 ディスプレイと少なくともキーボードを含む入力装置とを含むコンピュータを利用して設計、実施及び評価されるコンピュータ適応型テストを制御するシステムであって、固有の特性を有しており出題されるべき複数の問題によって構成されるデータベースと、第1番目から第 $n-1$ 番目までの問題に受験者が解答を与えた時点で前記第1番目から第 $n-1$ 番目までの問題に対する前記受験者の成績に基づき前記受験者の能力を推定する能力推定手段と、第 $n$ 番目の問題が出題される際に前記第1番目から第 $n-1$ 番目までの問題に対する前記受験者の成績に応答して前記データベースから第 $n$ 番目に出題すべき問題を選択する問題選択手段と、を備えたシステムにおいて、

前記能力推定手段によって推定される受験者の能力は、 $r_{ij}$ は受験者 $i$ が問題 $j$ に対して獲得する部分得点であり、 $P(\theta_i)$ は受験者 $i$ が問題 $j$ に正答する確率であり、 $Q(\theta_i)$ は $1-P(\theta_i)$ である場合に、

【数1】

$$L_{\text{part}}(\theta_i) = \prod_j^n P_j^{r_{ij}}(\theta_i) Q_j^{1-r_{ij}}(\theta_i)$$

によって表される部分得点モデルに基づく尤度 $L_{\text{part}}(\theta)$ を最大化する $\theta$ によって与えられることを特徴とするシステム。

【請求項2】 請求項1記載のシステムにおいて、前記 $P(\theta_i)$ は、2パラメータ・ロジスティック・モデルを用いて、

【数2】

$$P_j(x_{ij}=1|\theta_i) = \frac{1}{1 + \exp(-Da_j(\theta_i - b_j))}$$

として表され、この数式における $a$ 及び $b$ は前記データベースに記憶されている

複数の問題の固有の特性である識別力と困難度であり、 $D$ は定数であることを特徴とするシステム。

【請求項3】 ディスプレイと少なくともキーボードを含む入力装置とを含むコンピュータを利用して設計、実施及び評価されるコンピュータ適応型テストを制御する方法であって、前記コンピュータは、固有の特性を有しており出題されるべき複数の問題によって構成されるデータベースと、第1番目から第 $n-1$ 番目までの問題に受験者が解答を与えた時点で前記第1番目から第 $n-1$ 番目までの問題に対する前記受験者の成績に基づき前記受験者の能力を推定する能力推定手段と、第 $n$ 番目の問題が出題される際に前記第1番目から第 $n-1$ 番目までの問題に対する前記受験者の成績に応答して前記データベースから第 $n$ 番目に出題すべき問題を選択する問題選択手段と、を備えている方法において、

前記能力推定手段によって推定される受験者の能力は、 $r_{ij}$ は受験者 $i$ が問題 $j$ に対して獲得する部分得点であり、 $P(\theta_i)$ は受験者 $i$ が問題 $j$ に正答する確率であり、 $Q(\theta_i)$ は $1-P(\theta_i)$ である場合に、

【数3】

$$L_{\text{part}}(\theta_i) = \prod_j^n P_j^{r_{ij}}(\theta_i) Q_j^{1-r_{ij}}(\theta_i)$$

によって表される部分得点モデルに基づく尤度 $L_{\text{part}}(\theta)$ を最大化する $\theta$ によって与えられることを特徴とする方法。

【請求項4】 請求項3記載の方法において、前記 $P(\theta_i)$ は、2パラメータ・ロジスティック・モデルを用いて、

【数4】

$$P_j(x_{ij}=1|\theta_i) = \frac{1}{1 + \exp(-Da_j(\theta_i - b_j))}$$

として表され、この数式における $a$ 及び $b$ は前記データベースに記憶されている複数の問題の固有の特性である識別力と困難度であり、 $D$ は定数であることを特徴とする方法。

【請求項5】 請求項3又は請求項4記載の方法を実行するコンピュータ・プログラムが記憶されているコンピュータ可読な記憶媒体。

【請求項6】 請求項3又は請求項4記載の方法を実行するコンピュータ・プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、テスト結果の処理システム及び方法に関する。更に詳しくは、本発明は、項目反応理論の中の部分得点モデルに依拠してテストを設計、実施及び評価するシステム及び方法であって、特に、解答が単なる正誤の二値だけでなく複数の段階を有する部分点となることも許容するテストを従来よりも簡略化された態様で設計、実施及び評価するシステム及び方法に関する。

【0002】

【従来の技術及び発明が解決しようとする課題】

今日、テストを設計しその結果を処理する際に用いられる理論として、古典的テスト理論と項目反応理論との2つが広く知られている。これらのテスト理論に関しては、例えば、渡部洋編著『心理統計の技法』（福村出版、2002年）の6章及び7章にそれぞれ一般的な解説が与えられている。なお、テスト理論におけるテストとは、学力テストだけでなく、心理学における性格検査や臨床検査なども含むのであるが、この出願では、そのような広い応用分野を視野に入れた抽象的な説明を行うのではなく、理解を容易にするために、特に外国語試験などの学力テストを念頭において具体的な説明を試みたい。なお、項目反応理論における「項目」とは、学力テストの場合であれば問題を意味する。従って、この出願で具体例に即した説明する際には、「項目」ではなく「問題」、「被験者」ではなく「受験者」など、学力テストにおける一般的な用語を用いる。しかし、本発明は、一般のテスト理論に関する汎用的なものであって、学力テストだけに限定されることはなく、テスト理論が一般的に適用される上記の分野にも適用が可能である。

【0003】

古典的テスト理論は、現在広く用いられているが、得られる結論が受験者の特性の統計的な分布に依存するという特徴を有する。従って、古典的テスト理論に依拠している場合には、特性が異なる受験者の集団に対して実施された複数のテストで得られた平均点や偏差値を比較することは、適切でない。

#### 【0004】

古典的テスト理論の短所を克服する理論として項目反応理論があり、現在では、この項目反応理論に基づいて設計され結果が処理される学力テストが多く存在する。例えば、最近出版された項目応答理論の教科書である豊田秀樹著『項目反応理論入門編』（朝倉書店、2002年）の第16頁には次のような記述がある。「有名な語学試験であるTOEFL…は1年間に何度も実施される。しかも世界中で実施される。同一の被験者が再受験する可能性があるから同じ項目の集まりである同一のテストは2度と使用できない。このため平均点や通過率は、テストごとにそれぞれ異なる。地域によって英語力にも差があるから特性値の分布も異なる。したがって、偏差値や正規得点や経験分布関数を使用して、受験結果を異なるテスト間で比較することは不可能である。…TOEFLの点（たとえば、500点、650点など）は、どこで受験したか、いつ受験したか、どの問題を解いたかに関係なく留学の可否の判断に利用される。つまり異質な受験者が、異なる項目を、異なる日時に、異なる場所で受験したにも関わらず、被験者は、統一された処遇を受けることができる。この処遇を可能にするテストを継続的に、前向きに運用するシステムを構築する数理モデルが、…項目反応モデルである。」

#### 【0005】

また、特開2002-006734号公報には、出題が予定される試験問題を予め樹状に配置しておき、受験者による解答の正否に応じて樹状に配置された経路に沿って問題を順に出題し、単に正解の数だけでなく、どのような経路で最終地点まで到達したかという途中経過を考慮して受験者の能力を推定するテストの方法及びシステムが開示されている。この公開特許公報にも、項目反応理論への言及が見られる。

#### 【0006】

しかし、従来型の理論に基づく項目反応モデルでは、2 値によって採点が可能な○×式の解答以外の、例えば、1 と 0 との間の部分点を許容せざるを得ないような形式の問題の処理が容易ではなかった。段階反応モデルなど、多値の結果も扱えるように構築された方法も古くから存在してはいるが、従来の方法は後述する短所を有している。従って、部分得点を許容しなければ的確な採点が困難な英作文やスピーチなどは、コンピュータ適応型テスト（CAT）に含まれないことが通常である。例えば、上述の特開 2002-006734 号公報に開示されている発明では、項目反応理論の利用が明言され、出題されるべき問題が予め樹状に配置されているが、受験者がある問題に正解したら右下、誤答したら左下に配置された問題に移動するというように、解答が正誤のいずれかの 2 値であることが予定されている。

#### 【0007】

本発明は、このような従来型の 2 値の解答を予想する CAT とは異なり、部分得点を与えるような採点を許容する CAT であって、しかも、従来の段階反応モデル等の複雑なモデルよりもはるかに容易に結果処理が容易なシステム及び方法を提供することを目的とする。

#### 【0008】

##### 【課題を解決するための手段】

本発明によると、ディスプレイとキーボードやマウスなどを含む入力装置とを含むコンピュータを利用して設計、実施及び評価されるコンピュータ適応型テストを制御するシステムであって、固有の特性を有しており出題されるべき複数の問題によって構成されるデータベースと、第 1 番目から第  $n-1$  番目までの問題に受験者が解答を与えた時点で前記第 1 番目から第  $n-1$  番目までの問題に対する前記受験者の成績に基づき前記受験者の能力を推定する能力推定手段と、第  $n$  番目の問題が出題される際に前記第 1 番目から第  $n-1$  番目までの問題に対する前記受験者の成績に応答して前記データベースから第  $n$  番目に出題すべき問題を選択する問題選択手段と、を備えたシステムが提供される。本発明によるこのシステムでは、前記能力推定手段によって推定される受験者の能力は、 $r_{ij}$  は受験者  $i$  が問題  $j$  に対して獲得する部分得点であり、 $P(\theta_j)$  は受験者  $i$  が問題  $j$



に正答する確率であり、 $Q(\theta_i)$  は  $1 - P(\theta_i)$  である場合に、

【0009】

【数5】

$$L_{\text{part}}(\theta_i) = \prod_j^n P_j^{r_{ij}}(\theta_i) Q_j^{1-r_{ij}}(\theta_i)$$

【0010】

によって表される部分得点モデルに基づく尤度  $L_{\text{part}}(\theta)$  を最大化する  $\theta$  によって与えられる。

更に、本発明によるシステムでは、前記  $P(\theta_i)$  は、2パラメータ・ロジスティック・モデルを用いて、

【0011】

【数6】

$$P_j(x_{ij}=1|\theta_i) = \frac{1}{1 + \exp(-Da_j(\theta_i - b_j))}$$

【0012】

として表され、この数式における  $a$  及び  $b$  は前記データベースに記憶されている複数の問題の固有の特性である識別力と困難度であり、 $D$  は定数である。

また、本発明によると、ディスプレイとキーボードやマウスなどを含む入力装置とを含むコンピュータを利用して設計、実施及び評価されるコンピュータ適応型テストを制御する方法であって、前記コンピュータは、固有の特性を有しており出題されるべき複数の問題によって構成されるデータベースと、第1番目から第  $n-1$  番目までの問題に受験者が解答を与えた時点で前記第1番目から第  $n-1$  番目までの問題に対する前記受験者の成績に基づき前記受験者の能力を推定する能力推定手段と、第  $n$  番目の問題が出題される際に前記第1番目から第  $n-1$  番目までの問題に対する前記受験者の成績に応答して前記データベースから第  $n$  番目に出題すべき問題を選択する問題選択手段と、を備えている方法が提供される。本発明によるこの方法では、前記能力推定手段によって推定される受験者の

能力は、 $r_{ij}$ は受験者  $i$  が問題  $j$  に対して獲得する部分得点であり、 $P(\theta_i)$  は受験者  $i$  が問題  $j$  に正答する確率であり、 $Q(\theta_i)$  は  $1 - P(\theta_i)$  である場合に、

【0013】

【数7】

$$L_{\text{part}}(\theta_i) = \prod_j^n p_j^{r_{ij}}(\theta_i) Q_j^{1-r_{ij}}(\theta_i)$$

【0014】

によって表される部分得点モデルに基づく尤度  $L_{\text{part}}(\theta)$  を最大化する  $\theta$  によって与えられる。

更に、本発明による方法では、システムの場合と同様に、前記  $P(\theta_i)$  は、2パラメータ・ロジスティック・モデルを用いて、

【0015】

【数8】

$$P_j(x_{ij}=1|\theta_i) = \frac{1}{1 + \exp(-Da_j(\theta_i - b_j))}$$

【0016】

として表され、この数式における  $a$  及び  $b$  は前記データベースに記憶されている複数の問題の固有の特性である識別力と困難度であり、 $D$  は定数である。

更に、本発明による方法を実行するコンピュータ・プログラムが記憶されているコンピュータ可読な記憶媒体も提供される。

【0017】

更に、本発明による方法を実行するコンピュータ・プログラムも提供される。

【0018】

【発明の実施の形態】

本発明の基礎として用いる部分得点モデルについて述べる前に、一般的な2値の項目反応モデルについて簡単に説明する。2値とは、解答が正答又は誤答の2

つの値だけを取るという意味である。このモデルでは、受験者が問題に正答する確率を、受験者の能力を表すパラメータと問題の特徴付けるパラメータとを用いて表現する。例えば、本発明では、各問題を2つのパラメータ（識別力  $a$  及び困難度  $b$ ）によって特徴付ける2母数（パラメータ）ロジスティック・モデルを用いるのであるが、この場合、能力  $\theta$  を有する受験者  $i$  が問題  $j$  に正答する確率は次のように書くことができる。

【0019】

【数9】

$$P_j(x_{ij}=1|\theta_i) = \frac{1}{1 + \exp(-Da_j(\theta_i - b_j))}$$

【0020】

ここで、 $x$  は、受験者  $i$  が問題  $j$  に正答するならば1であり、誤答するならば0であるダミー変数である。また、 $D$  は定数である。この場合に、 $n$  問の問題を終了した時点での受験者の能力  $\theta$  の尤度  $L_B(\theta)$  は、次のように書くことができる。

【0021】

【数10】

$$L_B(\theta) = \prod_j P_j^{x_j}(\theta) Q_j^{1-x_j}(\theta)$$

【0022】

ここで、 $P(\theta)$  は数式9の右辺の正答確率であり、 $Q(\theta)$  は誤答確率つまり  $1 - P(\theta)$  である。項目応答理論では、数式10の尤度  $L_B(\theta)$  の最大値を与える  $\theta$  の値を受験者の能力パラメータの推定値とする最尤推定法が知られ、広く用いられている。ただし、数式10の右辺は積の形式で書かれており最大値を求めるのが容易でないので、和の形式に直して考察するため、両辺の自然対数を取った対数尤度  $\ln(L_B(\theta))$  の最大値を求めるのが一般的である。これは、自然対数は単調増加関数であって、尤度  $L_B(\theta)$  の最大値を与える  $\theta$  とその

自然対数  $\ln(L_B(\theta))$  の最大値を与える  $\theta$  とが一致するからである。

【0023】

以上が、従来から公知である一般的な2母数ロジスティック・モデルであるが、本発明において用いる部分得点モデルでは、問題への反応（解答）の評価が正誤という2値にとどまらず、部分点として評価されることが可能となる。具体的には、数式10のダミー変数  $x$  が1及び0の2値だけでなく、0から1までの3つ以上の複数の値（例えば、0、0.2、0.4、0.6、0.8、1の5つの値）を取ることを許容する。受験者  $i$  の問題  $j$  に対する部分得点を  $r_{ij}$  とすると、部分得点に対応する尤度は、次のように表すことができる。

【0024】

【数11】

$$L_{\text{part}}(\theta_i) = \prod p_j^{r_{ij}}(\theta_i) Q_j^{1-r_{ij}}(\theta_i)$$

【0025】

具体的な外国語試験、例えば英語の試験を考えると、正誤問題や多肢選択式問題であれば、正答及び誤答の2値として処理できる。しかし、英作文や英語での発話（質問に対する口頭での応答）は単純な正答又は誤答として評価するのは困難であり、部分点を与えることによって評価することが必要となる。従って、従来型の2値の項目応答モデルを用いることはできない。しかし、数式11によれば、任意の形式を有するテストの採点結果を部分得点に変換して分析用のデータとすることができる。

【0026】

数式11の意味を解釈するために、同一の問題パラメータを有する  $s_j$  個の問題が同じ受験者に向けて出題されたとする。学力テストの場合には、パラメータは同じであるが内容としては別の問題を提示すると想定するのが無理のない想定である。性格検査などに用いられる質問紙では、内容的にも同一の問題を提示するケースも考えられるが、これは項目反応理論の前提である局所独立の仮定に反する可能性も生じるので、ここでは、一応同一パラメータであるが内容の異なる質問項目を提示すると想定する。

【0027】

一般の項目反応モデルでは、このような想定条件下の尤度は、数式10から次のように表すことができる。

【0028】

【数12】

$$L_B(\theta_i) = \prod_j^n \prod_k^{s_j} P_j^{x_{ijk}}(\theta_i) Q_j^{1-x_{ijk}}(\theta_i)$$

$$= \prod_j^n P_j^{\sum_k^{s_j} x_{ijk}}(\theta_i) Q_j^{\sum_k^{s_j} (1-x_{ijk})}(\theta_i)$$

【0029】

すべての問題に関して繰り返しの解答  $s_j$  が等しい ( $=s$ ) と仮定した場合、この数式12の  $s$  乗根をとると、次の数式が得られる。

【0030】

【数13】

$$L_B^*(\theta_i) = \sqrt[s]{L_B(\theta_i)}$$

$$= \sqrt[s]{\prod_j^n P_j^{\sum_k^s x_{ijk}}(\theta_i) Q_j^{\sum_k^s (1-x_{ijk})}(\theta_i)}$$

$$= \prod_j^n P_j^{\frac{\sum_k^s x_{ijk}}{s}}(\theta_i) Q_j^{1-\frac{\sum_k^s x_{ijk}}{s}}(\theta_i)$$

【0031】

$L_B(\theta)$  と  $L_B(\theta)^*$  との最大値を与える  $\theta$  は同一であるので、最尤推定値は、数式12及び数式13で同じである。数式13において

【0032】

【数14】

$$\frac{\sum_k^s x_{ijk}}{s} = r_j$$

【0033】

とおけば、 $L_{\text{part}}(\theta)$  と  $L_B^*(\theta)$  とは形式上同じとなる。つまり、パラメータが同じである問題の繰り返しによる正誤の平均によって部分得点が表現される場合には、本発明が依拠する部分得点モデル  $L_{\text{part}}(\theta)$  の解と一般の項目反応理論による  $L_B(\theta)$  の解とは、と  $L_B^*(\theta)$  を通じて一致する。以上により、同一パラメータの問題を複数回実施し、その正誤の合計又は平均により成績をつけるようなタイプの部分得点に関しては、当該項目のすべての実施結果に関してその平均を取ることににより 0 から 1 までの間の部分得点を作ることになれば、部分得点モデル  $L_{\text{part}}(\theta)$  によって処理しても、従来の 2 値データによる処理と同一の結果となることがわかる。数式 14 から、提示の回数  $s_j$  を増加させれば、実質的に 0 から 1 までの任意の部分得点を表現できることが導かれる。なお、数式 12 と数式 13 との間で  $s$  乗根分の相違があると実際には不便なので、次のように、数式 11 を  $s$  乗して、その自然対数を部分得点化の対数尤度とするのが好ましい。

【0034】

【数15】

$$l_{\text{part}} = \sum_{j=1}^n s \{ r_j \ln(P_j(\theta)) + (1-r_j) \ln(Q_j(\theta)) \}$$

【0035】

さて、実際の学力テストでは、相互に関連する複数の問題が出題されることがある。外国語テストの例では、一定の長さの文章に関して複数の問題が出題される場合などがこれに該当する。しかし、このような問題群の存在は、項目反応理論の局所独立の仮定に反する可能性がある。

【0036】

因子分析的に考えてみると、仮にテストの得点  $z$  が、テスト問題の全体で測定する因子  $f_1$  と、ある問題群  $J$  に共通する因子  $f_2$  と、各問題に独自の因子  $u$  から構成されるとすれば、受験者  $i$  による問題  $j$  の得点は次のように表現することができる。

【0037】

【数16】

$$z_{ij} = f_{i1}a_{1j} + f_{i2}a_{2j} + u_j$$

【0038】

ここで、 $a$  は因子負荷である。問題  $j$  が問題群  $J$  に属する場合には  $a_{2j}$  はゼロでなく、問題  $j$  が問題群  $J$  に属する場合には  $a_{2j} = 0$  である。数式 16 において問題群  $J$  についての和を取ると、次のようになる。

【0039】

【数17】

$$\sum_J z_{ij} = \sum_J f_{i1}a_{1j} + \sum_J f_{i2}a_{2j} + \sum_J u_j$$

【0040】

右辺第2項は問題群  $J$  に共通する因子  $f_2$  に関する和であるが、因子の直交性の仮定からは、他の問題との相関を持たず独自性部分と見ることができる。項目反応理論でも項目に固有な成分を想定しているわけであるから、右辺第2項はこれに該当する。すなわち、局所独立の仮定に触れずに済むことになる。因子  $f_1$  を項目反応理論の能力パラメータ  $\theta$  と仮定すれば、テストで相互に関連する問題群の和を取って部分得点として処理することには不都合はない。局所従属の関係にある問題に関して本発明が提唱するような処理を行うことは、項目反応理論の仮定からはむしろ望ましいとさえいえる。

【0041】

もし完全な同一パラメータの項目でなく類似パラメータの項目の繰り返しを想

定する場合は、一般の2値の尤度は項目  $j$  の繰り返し提示の正誤パターンによって異なり、次のようになる。

【0042】

【数18】

$$L_B(\theta_i) = \prod_j^n \prod_k^{s_j} P_{ijk}^{x_{ijk}}(\theta) Q_{ijk}^{1-x_{ijk}}(\theta)$$

【0043】

$\theta$  固定、項目パラメータ所与とする。数式18の対数尤度を取り、類似項目  $j$  に係わる部分だけを取り出すと

【0044】

【数19】

$$l_{B(j)} = \sum_{k=1}^s (x_k \ln(P_k) + (1-x_k) \ln(1-P_k))$$

【0045】

が得られる（ただし正答確率などに対する添え字  $j$ 、及び  $\theta$  は省略する）。すると、その正誤反応  $X$  に関する期待値は

【0046】

【数20】

$$E(l_{B(j)}) = \sum_{k=1}^s (P_k \ln(P_k) + (1-P_k) \ln(1-P_k))$$

【0047】

となる。また

【0048】



【数 2 1】

$$\bar{P} = \frac{1}{s} \sum_{k=1}^s P_k$$

【0049】

【数 2 2】

$$P_k - \bar{P} = \Delta P_k$$

【0050】

とするとき、 $\Delta P_k$ は、各項目が類似項目であるため微小と仮定する。 $\Delta P_k$ の2次以上の項を無視すると

【0051】

【数 2 3】

$$\begin{aligned} P_k \ln(P_k) &= (\bar{P} + \Delta P_k) \ln(\bar{P} + \Delta P_k) \\ &\cong (\bar{P} + \Delta P_k) \left( \ln(\bar{P}) + \Delta P_k \times \frac{1}{\bar{P}} \right) \\ &\cong \bar{P} \ln(\bar{P}) + \Delta P_k \ln(\bar{P}) + \Delta P_k \end{aligned}$$

【0052】

【数 2 4】

$$\begin{aligned}(1-P_k)\ln(1-P_k) &= (1-\bar{P}-\Delta P_k)\ln(1-\bar{P}-\Delta P_k) \\ &\cong (1-\bar{P}-\Delta P_k)\left\{\ln(1-\bar{P}) - \Delta P_k \times \frac{1}{1-\bar{P}}\right\} \\ &\cong (1-\bar{P})\ln(1-\bar{P}) - \Delta P_k \\ &\quad - \Delta P_k \ln(1-\bar{P})\end{aligned}$$

【0053】

が成立する。数式 2 3 及び 2 4 を数式 2 0 に代入し、

【0054】

【数 2 5】

$$\sum_{k=1}^S \Delta P_k = 0$$

【0055】

であることを考慮に入れば

【0056】

【数26】

$$\begin{aligned}
 E(l_{B(j)}) &\cong \sum_{k=1}^s \{ \bar{P} \ln(\bar{P}) + (1-\bar{P}) \ln(1-\bar{P}) \} \\
 &\quad + \ln\left(\frac{\bar{P}}{1-\bar{P}}\right) \sum_{k=1}^s \Delta P_k \\
 &= \sum_{k=1}^s \{ \bar{P} \ln(\bar{P}) + (1-\bar{P}) \ln(1-\bar{P}) \} \\
 &= s \{ \bar{P} \ln(\bar{P}) + (1-\bar{P}) \ln(1-\bar{P}) \}
 \end{aligned}$$

【0057】

次に部分得点についても検討してみよう。先程と同様に全ての項目で繰り返しの回答  $s_j$  が等しい ( $=s$ ) と仮定した場合、(16) 式中の正誤の確率を (22) 式の平均反応確率を

【0058】

【数27】

$$\bar{Q}_j(\theta) = 1 - \bar{P}_j(\theta)$$

【0059】

で置き換える。すると

【0060】

【数28】

$$l_{\text{part}} = \sum_{j=1}^n s \{ r_j \ln(\bar{P}_j(\theta)) + (1-r_j) \ln(\bar{Q}_j(\theta)) \}$$

【0061】

となる。上式より特定の類似問題  $j$  に係わる部分だけを取り出し、また部分得点

$r$  が類似問題  $j$  の正誤の平均であることから

【0062】

【数29】

$$l_{\text{part}(j)} = s \left\{ \frac{\sum_{k=1}^s x_k}{s} \ln(\bar{p}) + \left(1 - \frac{\sum_{k=1}^s x_k}{s}\right) \ln(1 - \bar{p}) \right\}$$

【0063】

となる（ただし正答確率などに対する添え字  $j$ 、及び  $\theta$  は省略する）。すると、その期待値は

【0064】

【数30】

$$\begin{aligned} E(l_{\text{part}(j)}) &= s \left\{ \frac{1}{s} \ln(\bar{p}) \times E\left(\sum_{k=1}^s x_k\right) + \ln(1 - \bar{p}) \times E\left(1 - \frac{\sum_{k=1}^s x_k}{s}\right) \right\} \\ &= s \left\{ \frac{1}{s} \ln(\bar{p}) \times \sum_{k=1}^s p_k + \ln(1 - \bar{p}) \times \left(1 - \frac{1}{s} \sum_{k=1}^s p_k\right) \right\} \\ &= s \left\{ \bar{p} \ln(\bar{p}) \times (1 - \bar{p}) \ln(1 - \bar{p}) \right\} \end{aligned}$$

【0065】

であり、数式26及び30から

【0066】

【数31】

$$E(l_{B(j)}) \cong E(l_{\text{part}(j)})$$

## 【0067】

が成立する。尤度を構成する全ての類似項目に数式 31 が成立すれば、部分得点の最尤解と、2 値データによる解は近似的に一致すると考えられる。

なお、以上で、本発明における部分得点モデルは、同一パラメータの問題又は類似パラメータにおいて繰り返し定義の回数が同一であれば、通常の 2 値の項目反応モデルと対応関係を有することを既に示した。しかし、現実の適用場面では、すべての問題の繰り返し回数が同一である、つまり、部分得点の段階数がどの問題についても同一であるとは限らない。例えば、学力テストにおいては、初めに 2 値的に採点される易しい項目があり、後半に多値的に採点される難しい項目が配置されることが多い。このような状況に対応するためには、数式 15 を次のように拡張する必要がある。

## 【0068】

【数 32】

$$l_{\text{part}} = \sum_{j=1}^n s_j (r_j \ln(P_j(\theta)) + (1-r_j) \ln(Q_j(\theta)))$$

## 【0069】

このような拡張型部分得点モデルでも、一般の 2 値モデルの最尤解と数式 20 の解とは近似的に一致する。

以上では、本発明によるコンピュータ適応型テスト設計及び処理システムにおいて用いられる部分得点モデルについて説明した。この部分得点モデルは、2 値の項目反応モデルを修正することによって得られている。従って、この部分得点モデルでは、推定すべきパラメータの数も 2 値モデルの場合と同じであり、問題の特性に特別な関心があるとき以外は、学力テストの設計及び結果処理の際に、従来試みられてきたような複雑な多値モデルを利用する必要性は少ないといえよう。

## 【0070】

本発明において用いられる部分得点モデルを従来から公知であった段階反応モ

デルなどと比較すると、次のようなことが明らかとなる。(1) 項目反応理論では、通常はデータが2値で1次元であることを要求しているが、部分得点モデルは、多値、多次元データに適用できる。(2) 部分得点モデルは(段階反応モデルや他の多値モデルに比較して)簡明なモデルで利用者に理解しやすい。(3) 部分得点モデルは、(段階反応モデルや他の多値モデルと違い)一般によく利用されている2母数ロジスティックモデルとシームレスなので、結果の解釈も容易であり、2値と多値の混在データの分析にも便利である。(4) 部分得点モデルは、モデルのパラメータの数が(段階反応モデルや)他のモデルに比較して少なく、推定に問題が生じない。(5) 部分得点モデルは、どのような解答(回答)結果でも0から1の部分得点に変換すれば適用できるので応用範囲が広い。(6) 部分得点モデルは、テストだけでなく質問紙データにも容易に適用できる。(7) 部分得点モデルは、解答(回答)の段階数が増えても(段階反応モデルや他のモデルのように)モデルのパラメータが増加しないので、採点結果の変更に対応が容易である。

#### 【0071】

また、本発明の発明者の1人(藤森)は、部分得点モデルを用いたシミュレーションを行ったが、その結果によると、(1) 段階反応モデルは、テストが少数項目であると識別力の推定にバイアスがかかるが、部分得点モデルではこの現象は生じない。(2) モデル比較の基準として知られるAIC(赤池情報量規準)では、段階反応モデルに比較して部分得点モデルが良い。(3) 正答数得点と能力推定値の順位相関は、部分得点モデルの方が段階反応モデルより高い(真の値との相関では両モデルともほぼ同等)。

#### 【0072】

次に、上述の部分得点モデルを用いて英語テストを設計、実施及び評価する本発明のシステムについて説明する。コンピュータ適応型テスト(CAT)は既に知られており、実施されている。CATは、受験者全員がテスト用紙に印刷された同じ問題を提示され、鉛筆などの筆記用具を用いて解答を行う伝統的なテストとは異なる。CATでは、受験者は、パーソナル・コンピュータなどのコンピュータのディスプレイに問題を提示され、マウスやキーボードなどの入力装置を用

いて解答を行う。受験者に与えられる  $n$  番目の問題は、その受験者がテスト開始から 1 問前の  $n-1$  番目までの問題に対して与えてきた成績の関数として与えられる。

#### 【0073】

本発明に基づく CAT では、テストの前半部分において、受験者の解答を  $\bigcirc \times$  の 2 値によって評価することができる問題が出題される。この段階では、例えば、 $n$  番目が終了した時点での受験者の能力パラメータは、上述の数式 10 において与えられている尤度  $L_B(\theta)$  又はその自然対数である  $\ln(L_B(\theta))$  を最大化する  $\theta$  の値として推定される。その  $\theta$  の値に基づいて、出題されるべき問題群によって構築されておりその中に含まれる問題の識別力及び困難度が推定されている問題データベースの中から、 $n+1$  番目の問題が選択され出題される。受験者が解答する問題が  $\bigcirc \times$  の 2 値によって評価されるものである限りは、従来型の 2 値モデルを用いてテストにおける受験者の能力パラメータの推定及び次問の出題が制御される。

#### 【0074】

しかし、テストの後半部分には、2 値によっては評価することができない英作文やスピーチに関する問題が出題される。英作文の問題では、受験者は、キーボードを用いて、英文をタイプインする形式で解答する。スピーチの問題では、受験者がパーソナル・コンピュータに備え付けられたマイクロフォンに向かって、何らかの質問に対する応答として、又は、自由な内容で英語を話し、その音声として発話された内容が評価される。これらの問題は、採点者によって、部分得点を許容する態様で採点される。採点者は、受験者が現在利用しているパーソナル・コンピュータとインターネットなどのネットワークを介して接続されている別の端末の前で待機し、リアルタイムで採点を行うことも可能ではあるが、一般的には、受験者の解答をいったん記憶した上で集合的に多数の解答を採点する。

#### 【0075】

本発明が依拠する部分得点モデルは、この後半部分での採点に利用される。既に述べたように、段階反応モデルなどを用いて部分得点进行处理することも可能ではあるが、上述の部分得点モデルを用いた処理には多くの長所がある。

## 【0076】

図1を参照すると、本発明によるシステムの一例の概要が示されている。受験者は、本発明によるシステムによるテスト（例えば、英語能力テスト）を実施する語学学校などに設置された受験者ユニット101を用いてテストを受験する。解答結果は、インターネットなどのネットワークを介して、採点者ユニット102に送られ、採点者が部分得点を許容しながら採点を行う。

## 【0077】

図2には、受験者ユニット101の概要が図解されている。受験者ユニット101は、一般的なパーソナル・コンピュータであるのが通常であり、受験者は、ROM205から読み出されディスプレイ208に表示される自分向けにカスタマイズされた問題に対して、解答をキーボードやマウスなどの入力装置207を用いて解答する。解答結果、特に、部分得点を許容する採点が必要となるライティングやスピーキングに関する解答結果は、通信インターフェース202とインターネットなどのネットワーク103とを介して、採点者ユニット102に送られる。

## 【0078】

図3には、採点者ユニット102の概要が図解されている。採点者ユニット102は、受験者ユニット101と同じように、一般的なパーソナル・コンピュータであるのが通常であり、インターネットなどのネットワーク103を介して受験者ユニット101から送られてくる解答結果をディスプレイ308に表示し、キーボードやマウスなどの入力装置407を用いて採点する。

## 【0079】

図4には、本発明によるシステムを用いたテスト実施の概要を示す流れ図が示されており、特に、本発明が依拠する部分得点に関するライティング及びスピーキングに関する受験及び採点のプロセスが示されている。最初に、ステップ401では、ライティング及びスピーキングの問題が準備される。次に、プレテストを実施し（ステップ402）、本発明が依拠する部分得点モデルをプレテストの結果に適用して上述した尤度の計算を行い、個々の問題に固有のパラメータ（識別力及び困難度）を推定する（ステップ403）。これは、等化と称されるプ



プロセスである。このプレテストの実施による等化のプロセスを経ることによって、それぞれの問題の評価項目（アイテムに相当する）の識別力及び困難度を予め標準化することができ、後に受験者の能力を推定する際の尤度計算に利用できるようになる。このように、プレテストを行い、部分得点モデルによって修正された項目反応理論を適用することで、従来困難であった主観的なライティング及びスピーキングの問題の困難度及び識別力を標準化することが可能になり、プレテストの受験者で構成されるある特定の母体を基準とした絶対評価軸に基づく受験者の能力推定が可能となることも本発明の顕著な効果である。

#### 【0080】

次に、本発明が依拠する部分得点モデルを用いたライティング及びスピーキングの採点例について簡単に説明する。もちろん、これは、一例であって、別の項目を設定することも可能である。

#### 【0081】

ライティングの問題がそれぞれ3問（A、B、C）出題される場合に、次の評価項目すべてに関して、0から1までの間の部分得点を用いて採点する。例えば、A問題については目標達成度（Goal Achievement）、B問題については目標達成度、文法（Grammar）及び語彙（Vocabulary）、C問題については目標達成度、文法、語彙、文体（Style）及び機構（Mechanics）から構成される合計で9項目の部分得点（0～1の間の値をとる数値）でライティングの能力を推定する。

#### 【0082】

スピーキングに関しても同様であるが、スピーキングに関しては6問（A、B-1、B-2、B-3、C、D）が出題されるとすると、A問題については発音（Pronunciation）、B-1、B-2及びB-3問題については目標達成度及び発音、C及びD問題については目標達成度、文法、語彙、流暢さ（Fluency）から構成される合計15項目の部分得点（0～1の間の値をとる数値）でライティングの能力を推定する。このように、ライティングとスピーキングとのテスト問題数は少なくとも、1つの問題において、複数の独立した観点の評価項目をアイテムとみなして、パラメータ推定の的確さを担保している。

#### 【0083】

また、ライティング及びスピーキングに関する問題をデータベースから選択する際には、リーディングの推定結果  $\theta$  から、最も適切な識別力及び困難度の評価項目を含むライティング問題を選択し、リスニングの推定結果  $\theta$  から最も適切なスピーキング問題を選ぶように設計することができる。これは、ライティングの能力はリーディングの能力と相関関係を有し、スピーキングの能力はリスニングの能力と相関関係を有するという前提に基づいている。

#### 【0084】

本発明のシステムを用いて上述のような処理を行うことによって、特定の能力を有していると推定される受験者にとって、解答するまでもなく容易な問題や、およそ解答が不可能なほど困難な問題を解答させられるという問題の発生が抑制され、受験者は、能力を的確に推定するために最適かつ最小の問題で構成されるテストを受けることが可能となる。

#### 【0085】

明細書の最後に、本発明の発明者の1人（藤森）による段階反応モデルと部分得点モデルとの比較に関する論文（『項目反応理論による多値データの分析について一段階反応モデルと部分得点モデルー』）を、文中で言及されている図面（Figure）や表（Table）を除いて、以下に引用する。なお、この論文は、この出願の出願時点では未発表である。また、この論文は、学力テストだけを念頭にいたものではなく、より広い応用についても言及しているので、「問題」ではなく「項目」、「受験者」ではなく「被験者」など、以上の用例と異なる箇所が含まれている。

#### 【0086】

##### 1. 研究の目的

2母数ロジスティックモデル(Birnbaum, 1968)のような一般の項目反応理論では、正答、誤答のような2値データを処理することが出来るだけであり、多値のデータを分析することは出来ない。しかし心理学一般の研究では質問紙データなどで多値データを利用することも多い。項目反応理論でも多値データが全く扱えないわけではなく、幾つかの分析モデルが開発されている。例えばSamejima(1969)の段階反応モデル(graded response model)、評定尺度モデル(rating

scale model; Andrich, 1978) や partial credit model (Masters, 1982) はこの種の代表的なものとされる。しかし現実問題としては分析するためのソフトウェア環境が十分整備されていないこともあってか、これらのモデルが広く一般に利用されているとは言い難い。

#### 【0087】

このような状況に鑑み、本研究では藤森(2001)の部分得点モデル(partial test score model)と段階反応モデルとの比較をシミュレーション及び質問紙データの分析結果をもとに行い、項目反応理論を利用した多値データに関する分析方法の検討を行うことにする。

#### 【0088】

多値データへ項目反応理論を適用するとしても、どのようなモデルが良いのか検討する必要がある。このためには(1)理論的側面の検討、モデル母数の再現性や能力母数の分布などの影響を検討するための(2)シミュレーションによる検討、そして現実のデータにどの程度うまくあてはまるのかという(3)実証データに基づく検討が求められるであろう。このため、本研究では(1)を1.3節で、(2)を2節で、そして(3)を3節で行うことにする。

#### 【0089】

##### 1. 1. 段階反応モデル

$m$ 個の順序性を持つ段階反応を許容するSamejima(1969)の段階反応モデルを初めに説明する。ここではテストのある一つの項目だけを考えることにする。被験者の回答の各段階に対応して決まるダミー変数を  $u$  とし、ある段階より低い段階に反応したとき 0、そうでないとき 1 となるダミー変数を  $X$  とする。能力  $\theta$  を持つ被験者が  $X=1$  となる確率を

#### 【0090】

##### 【数33】

$$P_u^*(\theta) = P(X=1|\theta)$$

#### 【0091】

で表すとき

【0092】

【数34】

$$P_u(\theta) = P_u^*(\theta) - P_{u+1}^*(\theta)$$

【0093】

【数35】

$$P_u(\theta) > 0$$

【0094】

が段階反応モデルである。ただし

【0095】

【数36】

$$P_0^*(\theta) = 1$$

【0096】

【数37】

$$P_m^*(\theta) = 0$$

【0097】

である。数式33は段階の間を確定する境界反応曲線であり、数式34は段階の反応確率を表現する段階反応曲線となる。前記の条件を満たす限り数式33の関数はどのようなものでも自由であるが、本研究では良く利用される2母数ロジスティックモデル

【0098】

【数38】

$$P_u^*(\theta) = \frac{1}{1 + \exp(-a(\theta - b_u))}$$

$$u = 1, \dots, m-1$$

## 【0099】

を仮定する。数式 35 式の条件のため、 $a$  は数式 38 の全ての曲線に共通した母数となり識別力と呼ばれる。 $b_u$  は各段階の閾値に関係した母数であり困難度と呼ばれる。すなわち段階反応モデルでは項目ごとに 1 つの識別力と各段階の閾値に対応する  $m-1$  個の困難度母数を持っていることになる。

## 【0100】

## 1. 2. 部分得点モデル

部分得点モデルでは部分得点を  $r$  ( $0 \sim 1$  の範囲) とするとき母数推定のための対数尤度は次式で表される。

## 【0101】

## 【数 39】

$$l = \sum_{i=1}^N \sum_{j=1}^m s_j \{ r_{ij} \ln(P_j(\theta)) + (1-r_{ij}) \ln(Q_{ij}(\theta)) \}$$

## 【0102】

ここで  $i$  は被験者、 $\theta$  はその特性値を表す母数、 $j$  は項目、 $s$  は 2 値項目換算の繰返し数、 $P$  は 2 母数ロジスティックモデル、 $Q=1-P$  である。

すなわち部分得点モデルは 2 母数ロジスティックモデルをその基礎に置いている。また同一あるいは類似項目母数を持つ項目を被験者に対し潜在的に繰返し実施することを想定している。このとき  $r$  は繰返し実施の正誤平均と考えることができる。このような部分得点モデルと繰返し実施を考えた 2 母数ロジスティックモデルの最尤解は一致する（類似母数の時は近似）ことが証明できる（藤森, 2002a）。

## 【0103】

## 1. 3. 理論的側面からの検討

部分得点モデルの特徴は、以下のようなものである（藤森, 2002b）。（1）項目母数は識別力と困難度で一般の利用者にも比較的理解しやすい。（2）モデル母数の数が他の多値モデルと比較して少なく推定上の問題がない。（3）どのような解答／回答結果でも値を  $0 \sim 1$  の部分得点に変換すれば適用できるの

で応用範囲が広い。(4) 質問紙の評定尺度にも容易に適用できる。(5) 解答の段階数が増えても母数が増加しないので、採点結果の変更や微調整にも対応が容易である。(6) 2母数ロジスティックモデルとシームレスなので識別力や困難度母数の解釈もそのまま2母数ロジスティックモデルと同様に行えるので、2値と多値の混在データなどでの運用に便利である。(7) testlet/組間などの部分得点化により2値項目からなるテストに組間が存在する場合、組間を1つの部分得点項目としてとしての運用することが可能になり、副次的にテスト全体での母数の数が減る。(8) 残差得点にモデルを再適用して多次元データへの対応もできる。

(7) (8) については本研究では取り上げていないが特徴としては大きなものである。

#### 【0104】

続いて段階反応モデルの特徴は、以下のようなものである。(1) 項目反応理論で多値データに対応したモデルとして著名なものである。(2) 発表されてから30年以上の時間がたち応用研究も報告されている(例えば野口(1999)など)。(3) 公開されている分析用ソフトウェアMULTILOGが存在する。(4) 各段階反応曲線が得られるため、各段階に対する反応にまで関心を持つ場合は便利である。例えば境界反応曲線の母数を利用すれば、段階ごとの反応が $\theta$ 上のどの水準で発生するのかについて予測も出来る。

#### 【0105】

両モデルを比較してみると、モデル化の考えの違いが明らかである。段階反応モデルは、モデルの母数の数が多いだけ、部分得点モデルに比較してデータとの当てはまりも良くなると考えられる。もちろんモデル母数の数が多ければ、母数の正確な推定にはデータ量が必要であるなどの問題が生じる危険性がある。逆に部分得点モデルは単純であるだけ、当てはまりは劣ると予想されるが推定値などの安定性は良いと考えられる。この点に関しては赤池情報量規準AICなどの観点の評価が必要となろう。

#### 【0106】

また段階反応モデルの(4)の点については、部分得点モデルでは、モデル上

これは困難である。行うとすれば、事後的には段階ごとに被験者の分布を作成して検討することになるだろう。項目母数は所与であってもテスト実施の事前であれば、被験者の能力分布について適当なる仮定を設けて、シミュレーションを行い段階ごとに被験者の分布を作成して検討することになるだろう。いずれにしても段階反応モデルより手間がかかることになる。

#### 【0107】

このような利点がある一方で段階反応モデルは（そして今まで提案された殆どの多値モデルも）、各段階にモデル母数を置くため段階の変化に柔軟性を欠く問題がある。たとえば、能力テストで教師が採点時に20点満点である項目を、0、5、10、15、20の粗い5段階で評価していたとしたら5段階の段階反応モデルでの分析となる。しかし、もし誤字脱字などである答案を1点減点することにしたら、たちまちモデルの母数の値だけでなく、母数の数そのものが変化してしまうという問題点がある。質問紙でも評定を5段階で行っている項目を4段階にしたら、段階反応モデルでは、それまでの項目母数をそのまま利用するわけにはいかなくなってしまうのである。

#### 【0108】

以上のように、どちらのモデルも一長一短はあるのだが、各段階や評定についてはそれほどの関心を持たず、項目の全体的な特性と能力母数の推定だけに関心がある場合には、部分得点モデルで十分と思われる。

#### 【0109】

### 2. シミュレーションによる検討

実際のテストや質問紙において回答がどのようなメカニズムで生起しているのかは正確に知ることは出来ない。部分得点モデルと段階反応モデルのどちらがより現実に近いかは現時点で判断することは難しい。このため両モデルの比較をシミュレーションで行うにしても、データの作成にあたっては、どちらかのモデルに従ってデータを作成するべきかを定めることは困難である。あるモデルを仮定してデータを作成し、当該モデルと別のモデルで分析しても、別のモデルは良い成績を発揮することは期待できず、比較もうまく行えないだろう。このため本研究では、部分得点モデルを前提としたシミュレーションと段階反応モデルを前提

としたシミュレーションの2つを行うことにした。各モデルは自己のモデルに従ったデータと他のモデルのデータの2つを分析することになる。もちろん実際のデータはこの2者以外のメカニズムによる可能性もあるわけであるが、両モデルの公平な比較を行うことはできる。すなわちデータがモデルに従っている場合は、モデルの再現性がシミュレーションによって検討でき、またデータがモデルに従っていない場合はモデルの頑健性が評価できる。

### 【0110】

#### 2. 1. 部分得点モデルを前提としたシミュレーションデータ

部分得点モデルを前提としたシミュレーションデータは、以下のようにして作成した。まず部分得点モデルの構成要素として2母数ロジスティックモデルを仮定する。この2母数ロジスティックモデルの母数の分布型を以下のように定める。識別力母数は、平均0.65、標準偏差0.25、下限0.3、上限2.0の切断正規分布、また困難度母数は、平均0、標準偏差0.5の正規分布に従うと仮定する。能力母数 $\theta$ は平均0、標準偏差1.0の正規分布に従うと仮定する。能力母数 $\theta$ を標準正規分布に従って作成し、2母数ロジスティックモデルから予想される正答確率を、範囲0～1の一樣乱数と比較し、前者が下回る場合被験者の反応を正答1、上回る場合誤答0とする。2母数ロジスティックモデルに従う、この2値データパターンを、被験者数500、項目数200として各10回繰り返し作成した(データ1～10)。ただし5項目ずつ同一母数としている。続いて、このデータの同一母数の5項目ずつの正誤の和の平均をとり0、0.2、0.4、0.6、0.8、1.0の5段階の値をとる部分得点データとした。すなわち部分得点データは、被験者数は2値データと同様に500であるが、項目数は40となる。同様にして、前記の項目母数を所与として新たに被験者の能力母数 $\theta$ を500人分追加してクロスバリデーション用のデータを作成した。

### 【0111】

#### 2. 2. 段階反応モデルを前提としたシミュレーションデータ

段階反応モデルを前提としたデータは以下のようにして作成した。段階反応モデルも、その構成要素として2母数ロジスティックモデルを仮定する。またその母数の分布形も0節と同様である。またデータの段階数としては1から5の値をと



る5段階を仮定する。このため、モデルより段階間の境界反応曲線は4つ必要となる。この曲線群を確定するため、まず識別力母数を1つ分布に従って発生させ、これを各境界反応曲線に共通する識別力とする。続いて4つの困難度母数を作成し、最も小さいものを選択して、段階1と2の境界反応曲線の困難度とする。以下同様にして困難度の小さなものから順にとり各境界反応曲線の困難度を決定する。これらの境界反応曲線間の差をとって各段階反応曲線とする。標準正規分布に従う能力母数 $\theta$ を1つ作成し、この値を固定して各段階反応曲線で予想される反応確率の区間（ $\theta$ を固定したときの全段階反応曲線の大きさの和は1である）に、0～1の一様乱数が入ったとき当該反応が生じたことにする。以上の過程を500人分繰返し母数推定用のデータとした。また0節と同様に、以上で確定した項目母数を利用してクロスバリデーション用のデータを500人分作成した。

#### 【0112】

### 2. 3. 母数の推定

母数の推定は、両モデルとも項目母数と潜在特性値 $\theta$ の交互同時最尤推定による自作のFORTRANプログラムによった。部分得点モデル用では交互同時最尤推定だけでなく項目母数の周辺最尤推定が可能であるが、段階反応モデルの推定プログラムは交互同時最尤推定のみに対応しているため、比較の便を考慮して両モデルとも項目母数を最尤推定した（結果は省くが部分得点モデルに関しては周辺最尤推定と交互同時推定の結果に大きな差はない）。両モデルとも $\theta$ と困難度の推定値は-3.5～3.5の範囲と定め、識別力の推定値の範囲は0.02～2.0としている。

#### 【0113】

### 2. 4. シミュレーションの結果と考察

表1は、部分得点モデルにより作成したシミュレーションデータ（以下部分得点データと呼ぶ）の能力母数の真値、正答数得点と両モデルで推定した $\theta$ の推定値の相関である。ただし相関は、ケンドールの順位相関係数を求めている（以下特に明示しない場合は相関はケンドールの順位相関を指す）。一般によく利用されるピアソンの積率相関でなくて順位相関を求めたのは、項目反応理論の現実の運

用場面では正答数得点などと推定値の順位の逆転現象が問題となるケースが多いためである。表1より明らかなように真の $\theta$ との相関は、いずれも部分得点モデルによる推定値の方が高いものの段階反応モデルとの差はほとんどない。これに比し、正答数との相関は、若干差が拡大し部分得点モデルによる推定値が高い相関を与えている。

#### 【0114】

なお部分得点モデルの項目母数の推定値に関しては、例えばデータセット1については識別力の平均自乗誤差(MSE)は、0.0014、困難度に関しては0.0017であり、特に問題はない精度で推定値が得られている。ちなみに、データ作成の元となったモデルとは異なるので値の良し悪しは判断できないが、段階反応モデルの識別力とのMSEは0.027、困難度は0.31となっている。ただし困難度に関しては、段階反応曲線の困難度の推定値平均によってMSEを算出している。

#### 【0115】

さて統計的な視点からはAIC (Akaike information criterion, 赤池情報量規準)

#### 【0116】

【数40】

$$AIC = -2 (\text{モデルの最大対数尤度}) \\ + 2 (\text{モデルの自由母数の数})$$

#### 【0117】

がモデルよさと自由母数の数のバランスをどのようにとるかについて我々に示唆を与えてくれる(赤池情報量規準については、詳しくは坂元ら(1983)など参照のこと)。AICでは値が小さいほど良いモデルとされるため、モデルの自由母数の数は抑制されることになる。試みにデータセット1の段階反応モデルのAICを求めたところ、46245.34であり、部分得点モデルは22142.46となっている。AICは小さい値であるほうが良いのであるから圧倒的に部分得点モデルが良いことになる。これはいずれのデータセットについても全く同様で

あり、個々の数値を問題にする必要は無いと判断し他の結果は省略する。なぜこのような圧倒的な差が生じるかと言えば、段階反応モデルの段階 2 ～ 4 の生起確率に小さくなるものがかなりあり、これが対数尤度の計算の際に影響しているためである。本研究では、AIC の観点からは明確な差が両モデルであったと考えるが、理論的側面の検討も含め将来的にはカルバック・ライブラー情報量による比較も行うことが適当であろう。

#### 【0118】

続いて部分得点データの推定値を利用してクロスバリデーションデータの成績を比較してみよう（表 2）。クロスバリデーションにおいても様相は先と同様で、真値との相関はいずれも部分得点データがほんの僅か上回り、正答数得点に関してはややモデル間の差が拡大することが示されている。

#### 【0119】

さて段階反応モデルにより作成したシミュレーションデータ（以下段階反応データと呼ぶ）についての結果が表 3 である。また同データに基づき推定された項目母数をクロスバリデーションデータに適用した結果が表 4 である。どちらにおいても部分得点モデルに比べて段階反応モデルの方が真の  $\theta$  との相関がほんのわずかに高いが、大きな差ではなく、データセット 1, 5 のように逆転しているケースも見受けられる。また全体的に真値との相関が表 1 及び表 2 に比較してやや低くなっており、段階反応モデルが部分得点モデルに比較して再現性が難しいモデルであることを示唆していると言えよう。段階反応データであるにもかかわらず、正答数得点との相関は段階反応モデルより部分得点モデルによる推定値の方が高くなっている。しかも部分得点データの場合よりその差が拡大していることが分る。合計点との順序関係の逆転は段階反応モデルに多くなっているのであるが、この逆転が生じる理由は、段階反応モデルでは項目によって評定の間部分の段階反応曲線が相対的にかなり低くなり、他項目の回答結果によっては、当該項目の回答が  $\theta$  の推定に（推定誤差の大きさはともかくとして、その点推定値の決定には）殆ど影響力を持てなくなるためである。この現象は段階反応モデルだけでなく、選択肢やカテゴリーごとの反応確率曲線をモデル化する他の多値モデルにも共通するため、これらのモデルの利用の際には注意を払う必要があると考

えられる。もちろん部分得点モデルも 2 母数ロジスティックモデルをベースとしている以上、合計点と  $\theta$  の逆転も一部生じざるを得ないが、その程度は低く抑えられている。

### 【0120】

次に段階反応モデルの項目母数の推定値について検討しよう。例えばデータセット 1 については識別力の平均自乗誤差 (MSE) は、0.2993、困難度に関しては 0.0636 であり、やや部分得点モデルの場合に比較して推定精度が悪いことがわかる。表 5 に示した項目 8, 15 はデータセット 1 の中の推定成績が悪い項目であるが、いずれも真の境界反応曲線が互いに接近しすぎている場合に推定が出来なくなっていることがわかる。このような場合には、2 つの境界反応曲線を 1 つとして処理するなどの対処が必要になると考えられるが、どのような基準で行うかなどの問題もあり本研究ではこれを行っていない。ある意味ではこのような対処を必要とすることに段階反応モデルの問題点が存在するとも言えよう。なおモデルは異なるものの部分得点モデルの識別力の MSE は 0.0136 となり段階反応モデルよりも良い推定値を与えている。ただし段階反応モデルの項目母数の推定に関しては周辺最尤推定では成績が改善する可能性も残されているので、ここではこれ以上取り上げず、別の研究で報告することにしたい。

### 【0121】

以上をまとめると本研究のシミュレーションの条件下では、部分得点データ、段階反応データのいずれであっても真値との相関は、どちらのモデルを利用しても大きな差は生じていない。これに対し正答数得点との相関は明らかに部分得点モデルの方が良い成績であり、特に予期に反し段階反応データで差が大きくなっている。また項目母数の数が多い段階反応モデルで危惧されたクロスバリデーションデータでの不適合は生じていない（ただし母数推定上の問題は残る）。これは一つには、本研究が特にノイズをのせていない単純なシミュレーションであることも影響しているかもしれない。

### 【0122】

## 3. 質問紙データによる検討

### 3. 1. 質問紙データ及び母数の推定

本研究で分析の対象としたデータは、〇県の青少年基本調査の中高生 1849 人のデータで、教師と生徒の心理的距離について調べた 9 項目尺度である（表 6）。各項目は「1. よくあてはまる」から「4. まったくあてはまらない」の 4 段階で評定を求めるものであるが、9 項目とも回答は値が小さいほど心理的距離が小さくなるように処理されている。部分得点モデルでは、このデータを 0～1 の部分得点として処理するために、各回答結果を  $x$  とするとき  $r = (x-1)/3$  と変換して「0, 1/3, 2/3, 1」の 4 段階をとる部分得点データとした。また推定方法はシミュレーションと同様の方法である。なお本調査の詳細については木原ら（1997）参照されたい。

### 【0123】

#### 3. 2. 質問紙データの結果と考察

段階反応モデルの識別力は部分得点モデルより全体的に高くなっている（表 7）。これと類似の現象はデータを 2 値化して通常の 2 母数ロジスティックモデルで分析した場合にも生じる。たとえば 4 段階評定の場合、2 値化は実質的には段階 2 を 1 に、段階 3 を 4 に置き換えるために生じる現象である。表 7 の「2 値モデル」はこのようなデータ変換をした結果である。部分得点モデルに比較してやや識別力が高くなっていることが分る。このような現象が発生することは、図 1 の項目 4 の群別の平均回答結果を 4 値と 2 値の場合で比較してみると良く理解できる。ただし図の作成にあたっては、4 値 2 値いずれの場合も全ての回答を 0～1 の範囲に変換している。多値データの 2 値化が識別力の推定値に大きな影響を与えることが図より理解できる。さて境界反応曲線が実質的には多値の回答を 2 値化して得られるデータに基づき定義され、推定されるため、この 2 値化と類似の現象が生じている可能性もある。しかしシミュレーションの結果では、部分得点モデルの識別力と段階反応モデルの識別力の推定値は大きく異なっていなかったのだから、2 値化を原因とすることにはやや疑問も残る。

### 【0124】

もう一つの可能性として、テスト項目数の違いがある。シミュレーションでは 40 項目であったが、本質問紙は 9 項目であり、この違いが影響した可能性がある。このため 1 つシミュレーションを追加した。段階反応データのデータセット

1の初めの10項目だけを取り、段階反応モデルにより交互同時推定値を求めたところ、識別力の値は表8のようになった。明らかに項目数が少なくなると識別力が大きくなる傾向があることが分る(MSE=0.2647)。段階反応モデルでは少数項目での識別力の推定にバイアスがかかることは明らかとしても、その原因がモデルに存在するのか、それとも推定方法に問題があるのかは区別されなければならない。つまり段階反応モデルそのものではなくて、同時最尤推定という方法が推定に悪影響を及ぼしている可能性もある(同時最尤推定による項目母数の推定値が一致性を持たないことは良く知られている)。このため、MULTILOG(Thissen, 1991)により周辺最尤推定値も算出したが、バイアスはほとんど変わらない(MSE=0.2558)。このことは推定法の問題というよりはモデルに問題があることを示唆するものであるかもしれない。なお表8からは項目7は一見すると同時最尤推定の識別力が極端に低く、反対にMULTILOGの推定値は妥当な値を得ているようであるが、実はMULTILOGは困難度において異常な推定値となっており、うまく推定できていないことには変わりはない。このためMSEの計算からは同項目を除外している。

#### 【0125】

部分得点データセット1の10項目を利用した部分得点モデルの分析では交互同時推定のMSEは、識別力が0.0094、困難度が0.0007であり、周辺最尤推定では識別力が0.0032、困難度が0.0014となった(表9)。同モデルでは交互同時推定、周辺最尤推定ともこのバイアスは認められないといっただろう。表10には質問紙データを両モデルで分析し、周辺最尤推定を利用した推定値を示してある。同じく表10より、シミュレーションと同様に、周辺最尤推定を利用しても段階反応モデルでは識別力が大きくなっていることが分る。

#### 【0126】

段階反応モデルにおいて少数項目では識別力の上方バイアスがかかることについては、慎重な検討を要するので、これについては別の研究で行うこととし、ここでは表7の同時最尤推定の結果に基づいて検討を進める。図2に項目1の部分得点モデルの項目反応曲線を、また図3に段階反応モデルの反応曲線を示した。

評定結果の合計点と両モデルの  $\theta$  のケンドールの順位相関を求めたところ段階反応モデルとは 0.9104、部分得点モデルとは 0.9424 の相関を得た(無回答の無いケース 1588 人のみで算出した。)。部分得点モデルの方が、合計点との相関がやや高く、その再現には適していることはシミュレーションの結果と一致している。相関の値は段階反応データのシミュレーション結果から得られた両モデルの相関と近いものとなっているが、これだけで本データは段階反応モデルに従って発生していると決めるわけには行かないだろう。その理由としては、たとえばシミュレーションデータの作成の際に多次元性などの影響を加味すれば、真値と推定値の相関だけでなく正答数(質問紙の場合は合計点)との相関も低下すると容易に予想できることなどがあげられるからである。

#### 【0127】

本節のデータは実際の調査であるため真値は不明であるからどちらのモデルが良いかを決定することは困難である。しかし段階反応モデルにおいて合計点との相関が低いことは、順位の逆転が十分説明がつくようなものであれば良いのであるが、一見で分るような性格のものではないため、モデルの利用者には逆転現象を納得しにくいということは利用の際には考慮しておいた方が良い。いずれにしろシミュレーションの結果からは、(シミュレーションの条件に近い場合は)真の  $\theta$  値の推定という点では大きな差は両モデルでないことが想像される。しかし、少数項目の場合の段階反応モデルの推定値のバイアスがどのような影響をもたらすのかは、もう少し検討する必要があるだろう。いずれにしろ項目特性として識別力や項目困難度以外に大きな関心がない場合、つまり回答の各段階の困難度などに特別な関心がない場合(質問紙を利用した多くの研究はこれに該当する)には、複雑で多くの母数を推定しなければならない段階反応モデルより簡単な部分得点モデルの方が適していると言えるだろう。もちろん、個々の段階の反応にまで関心を持つ場合はこの限りではない。

#### 【0128】

文献

Andrich, D 1978 A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.

Birnbaum, A. 1968 Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), Statistical theories of mental test scores (pp.395-479). Reading, MA: Addison-Wesley.

木原孝博・田中治彦・藤森進 1997 中学生・高校生の学校不適應に関する研究—岡山県青少年基本調査(1994)に基づく分析— 岡山大学教育学部研究集録, 104, 105-122.

藤森進 2001 項目反応理論における部分得点の処理について 日本教育心理学会第43回総会発表論文集, 394.

藤森進 2002a 項目反応理論におけるテストの部分得点の処理方法について 未発表論文.

藤森進 2002b 部分得点モデルとその応用 第1回心理測定研究会.

Masters, G.N. 1982 A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.

野口裕之 1999 適応型テストへの応用: CAT方式による識別性検査 渡辺直登・野口裕之編著「組織心理測定論」第8章 白桃書房.

坂元慶行・石黒真木夫・北川源四郎 情報量統計学 共立出版.

Samejima, F. 1969 Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph, No.17.

Thissen, D 1991 Multilog user's guide. Chicago, IL: Scientific Software.

#### 【図面の簡単な説明】

##### 【図1】

本発明によるシステムの一例の概要が示されている。

##### 【図2】

受験者ユニットの概要が図解されている。

##### 【図3】

採点者ユニットの概要が図解されている。

##### 【図4】

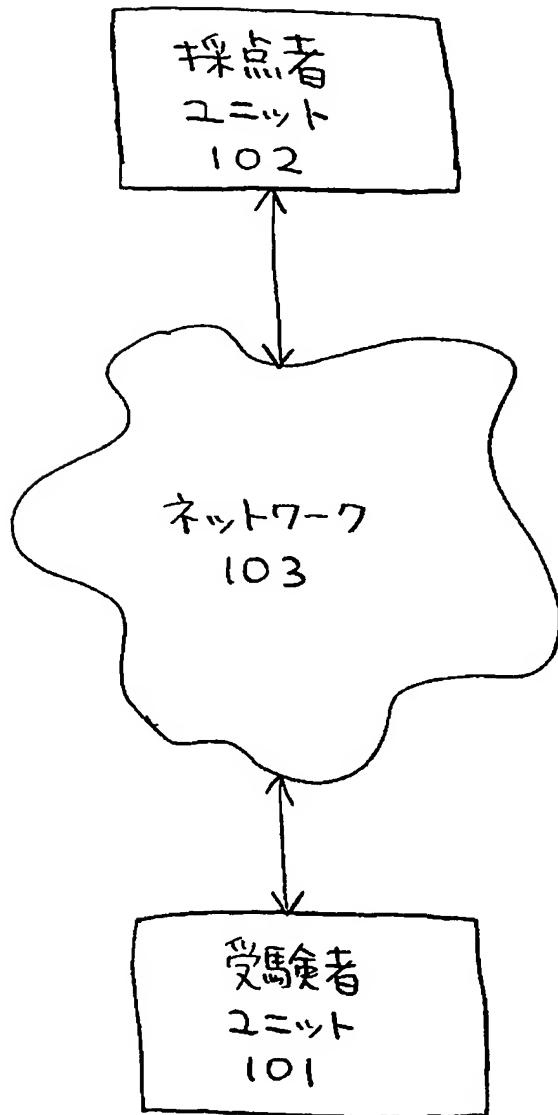
本発明によるシステムを用いたテスト実施の概要を示す流れ図であり、特に、本発明が依拠する部分得点に関するライティング及びスピーキングに関する受



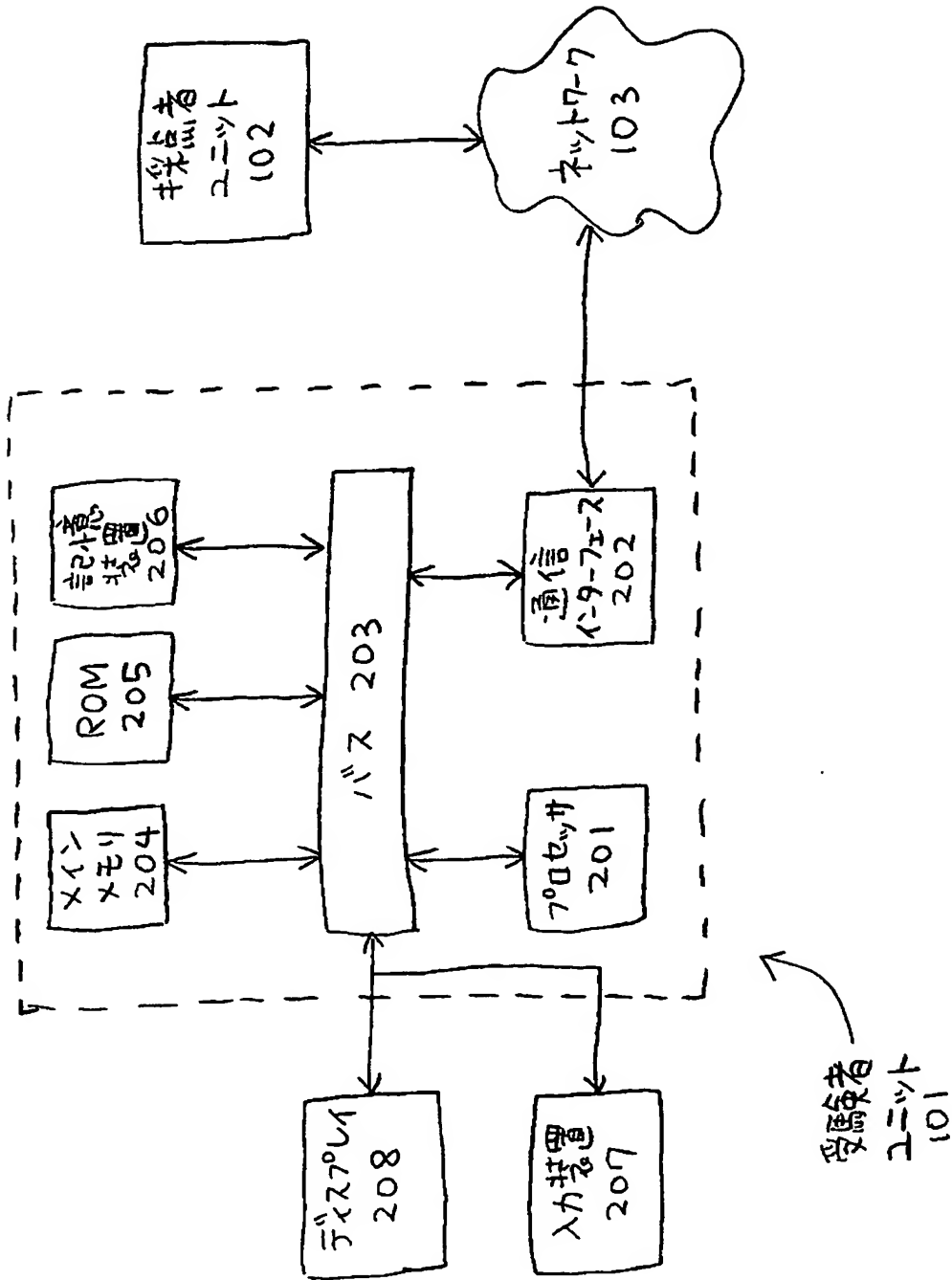
験及び採点のプロセスが示されている。

【書類名】 図面

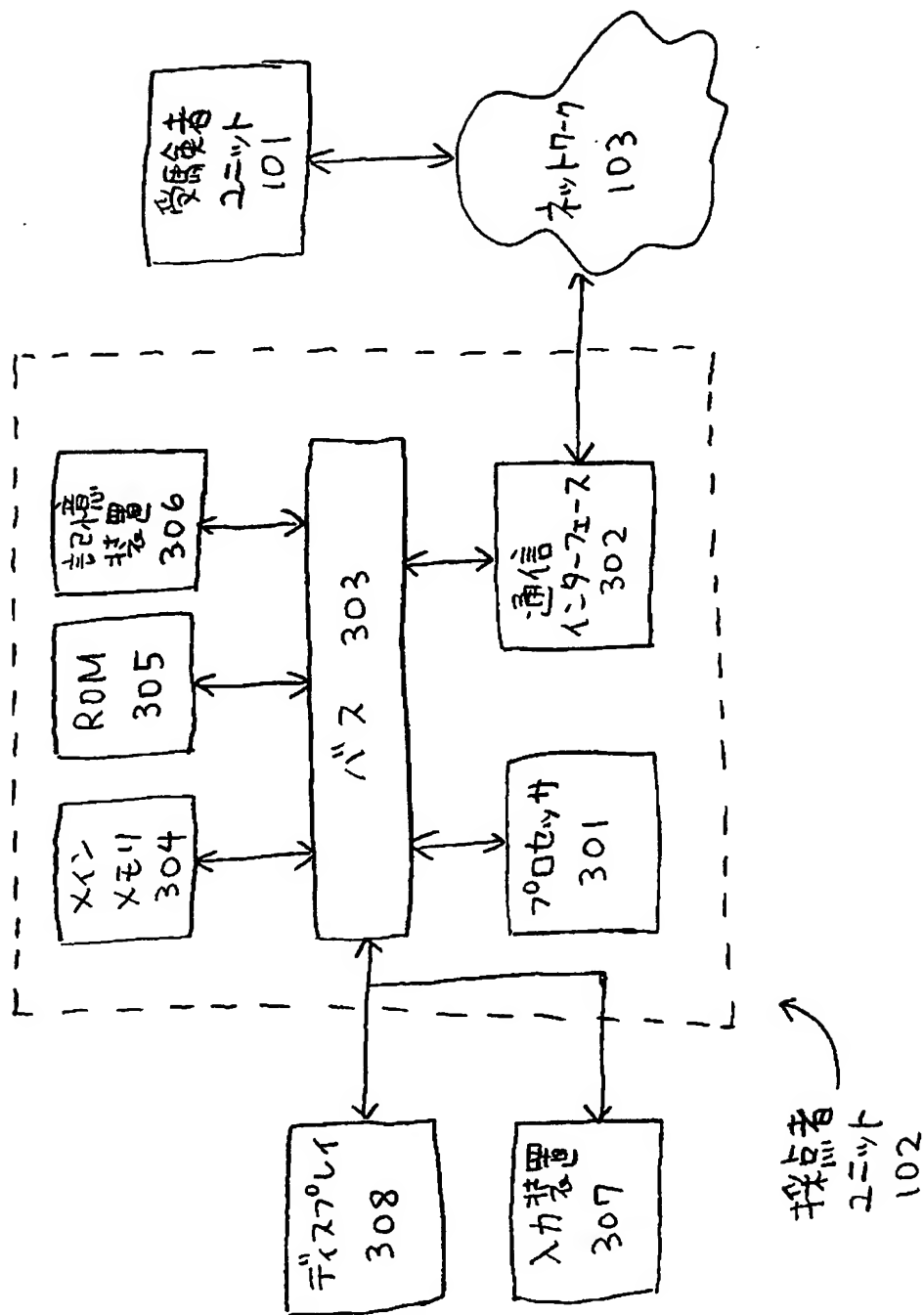
【図1】



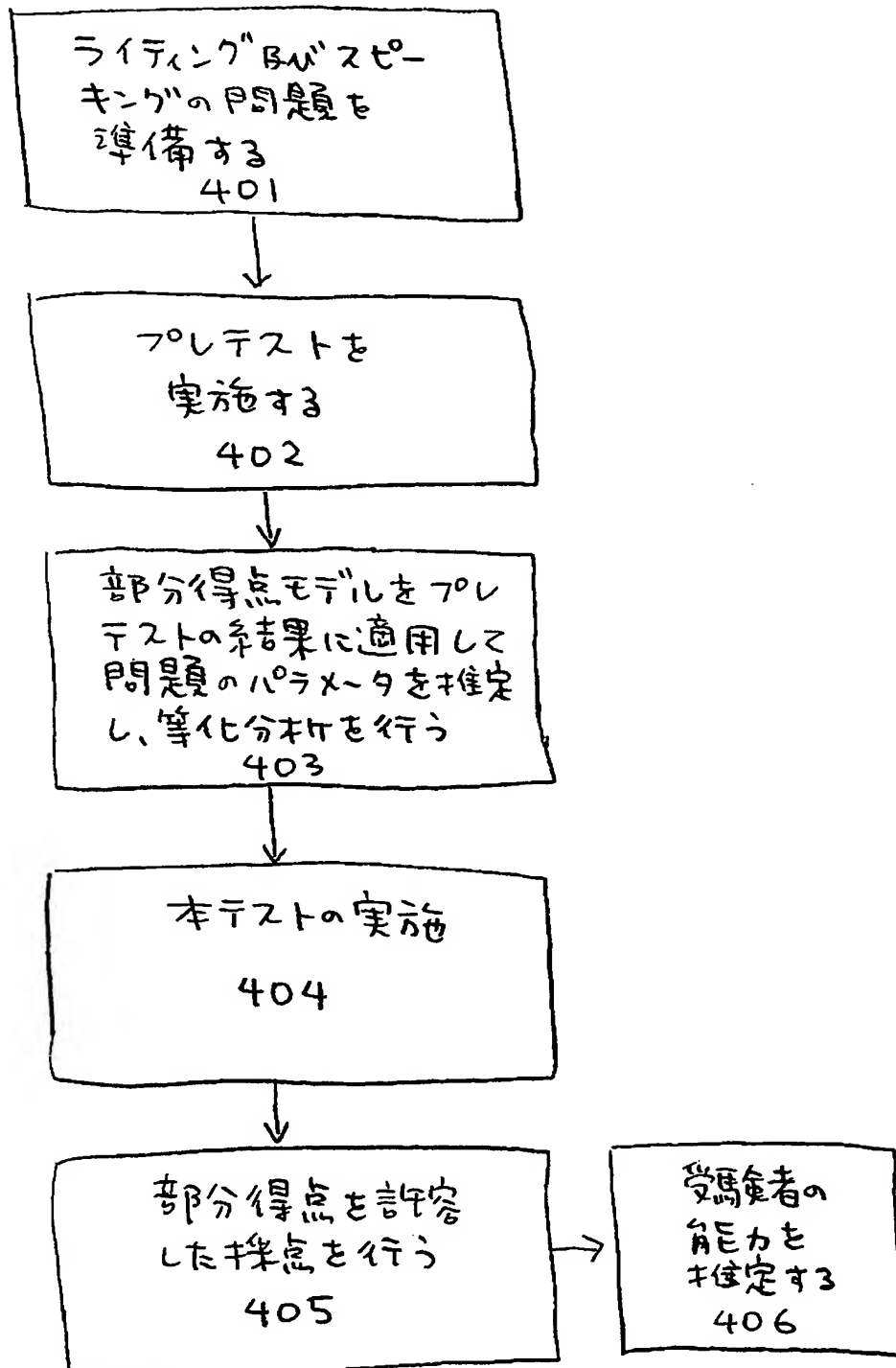
【図 2】



【図3】



【図4】



【書類名】 要約書

【要約】

【課題】 部分得点を与えるような採点を従来よりも容易に許容するコンピュータ適応型テスト（CAT）を実現するシステムを提供する。

【解決手段】 入出力装置（207、208）を含むコンピュータを利用して設計、実施及び評価されるコンピュータ適応型テストを制御するシステムが提供される。テスト結果に基づき受験者の能力の推定する際には、従来の項目反応理論を修正した部分得点モデルを用いることにより、外国語テストにおけるライティングやスピーキングのように正誤だけでは判定できない部分得点を要するような出題形式のテストの制御が可能となる。また、プレテストを行い（402）部分得点モデルを適用する（405）ことで、問題の困難度及び識別力を標準化することが可能になる。

【選択図】 図4

特願 2 0 0 2 - 2 7 9 5 1 3

出 願 人 履 歴 情 報

識別番号

[ 5 9 1 1 2 1 4 9 8 ]

1. 変更年月日

1 9 9 6 年 1 2 月 2 5 日

[変更理由]

名称変更

住 所

岡山県岡山市南方 3 - 7 - 1 7

氏 名

株式会社ベネッセコーポレーション

特願 2002-279513

出願人履歴情報

識別番号

[502348154]

1. 変更年月日  
[変更理由]

2002年 9月25日

新規登録

住所  
氏名

東京都府中市武蔵台2-36-10

藤森 進